



**Homologous proteins:**

**Top 10 BLAST Hits**

	Score	E
CRA 103000001517087 /altid=gi 10946770 /def=ref NP_067386.1  RA...	425	e-117
CRA 1000682330460 /altid=gi 7657492 /def=ref NP_055168.1  RAB26...	297	4e-79
CRA 18000004977238 /altid=gi 1710022 /def=sp P51156 RB26_RAT RA...	294	3e-78
CRA 18000005013109 /altid=gi 1083775 /def=pir  JC2528 GTP-bindi...	293	7e-78
CRA 89000000198627 /altid=gi 7296421 /def=gb AAF51708.1  (AE003...	273	9e-72
CRA 18000005076419 /altid=gi 7438397 /def=pir  T15123 hypotheti...	207	4e-52
CRA 18000004912300 /altid=gi 134236 /def=sp P20791 SAS2_DICDI G...	203	7e-51
CRA 98000043536338 /altid=gi 12963499 /def=ref NP_075615.1  cel...	203	9e-51
CRA 18000004929618 /altid=gi 131798 /def=sp P24407 RAB8_HUMAN R...	202	1e-50
CRA 18000004952869 /altid=gi 131848 /def=sp P22128 RAB8_DISOM R...	202	2e-50
CRA 18000005221564 /altid=gi 4586580 /def=dbj BAA76422.1  (AB02...	202	2e-50

**BLAST dbEST hits:**

	Score	E
gi 13033710 /dataset=dbest /taxon=960...	1318	0.0
gi 12785775 /dataset=dbest /taxon=960...	1316	0.0
gi 12904236 /dataset=dbest /taxon=960...	1035	0.0
gi 9093496 /dataset=dbest /taxon=9606...	694	0.0

**EXPRESSION INFORMATION FOR MODULATORY USE:**

library source:

From BLAST dbEST hits:

gi|13033710 prostate  
gi|12785775 brain  
gi|12904236 T cells from T cell leukemia  
gi|9093496 leukopheresis

From tissue screening panels:

leukocyte

FIGURE 1, page 2 of 2

1 MTGTPGAVAT RDGEAPERSP PCSPSYDLTG KVMLLGDTGV GKTCFLIQFK  
 51 DGAFLSGTFI ATVGIDFRNK VVTVDGVRVK LQIWDTAGQE RFRSVTHAYY  
 101 RDAQALLLLY DITNKSSFDN IRAWLTEIHE YAQRDVVIML LGNKADMSSE  
 151 RVIRSEDGET LAREYGVVFL ETSAKTGMNV ELAFLAIAKE LKYRAGHQAD  
 201 EPSFQIRDYV ESQKKRSSCC SFM (SEQ ID NO:2)

#### FEATURES:

##### Functional domains and key regions:

[1] PDOC00001 PS00001 ASN\_GLYCOSYLATION  
 N-glycosylation site

114-117 NKSS

[2] PDOC00004 PS00004 CAMP\_PHOSPHO\_SITE  
 cAMP- and cGMP-dependent protein kinase phosphorylation site

Number of matches: 2

1 214-217 KKRS  
 2 215-218 KRSS

[3] PDOC00005 PS00005 PKC\_PHOSPHO\_SITE  
 Protein kinase C phosphorylation site

Number of matches: 5

1 29-31 TGK  
 2 113-115 TNK  
 3 149-151 SER  
 4 173-175 SAK  
 5 212-214 SQK

[4] PDOC00006 PS00006 CK2\_PHOSPHO\_SITE  
 Casein kinase II phosphorylation site

116-119 SSFD

[5] PDOC00008 PS00008 MYRISTYL  
 N-myristoylation site

Number of matches: 5

1 3-8 GTPGAV  
 2 6-11 GAVATR  
 3 39-44 GVGKTC  
 4 52-57 GAFLSG  
 5 57-62 GTFIAT

[6] PDOC00017 PS00017 ATP\_GTP\_A  
 ATP/GTP-binding site motif A (P-loop)

36-43 GDTGVGKT

[7] PDOC00579 PS00675 SIGMA54\_INTERACT\_1  
 Sigma-54 interaction domain ATP-binding region A signature

32-45 VMLLGDTGVGKTCTF

#### Membrane spanning structure and domains:

Helix	Begin	End	Score	Certainty
1	48	68	0.715	Putative

# **BLAST Alignment to Top Hit:**

```
>CRA|103000001517087 /altid=gi|10946770 /def=ref|NP_067386.1| RAB37,
      member of RAS oncogene family; GTPase Rab37 [Mus
      musculus] /org=Mus musculus /taxon=10090 /dataset=nraa
      /length=223
      Length = 223
```

```
Score = 425 bits (1081), Expect = e-117
Identities = 209/223 (93%), Positives = 215/223 (95%)
Frame = +3
```

```
Query: 42  MTGTPGAVATRDGEAPERSPPCSPSYDLTGKVMLLGDTGVGKTCFLIQFKDGAFLSGTFI 221
          MTGTPGA      DGEAPERSPP SP+YDLTGKVMLLGD+GVGKTCFLIQFKDGAFLSGTFI
Sbjct: 1  MTGTPGAATAGDGEAPERSPPFSPNYDLTGKVMLLGDSGVGKTCFLIQFKDGAFLSGTFI 60
```

```
Query: 222 ATVGIDFRNKVVTVDGVVRVKLQIWDTAGQERFRSVTHAYYRDAQALLLLYDITNKSSFDN 401
          ATVGIDFRNKVVTVDG RVKLQIWDTAGQERFRSVTHAYYRDAQALLLLYDITN+SSFDN
Sbjct: 61  ATVGIDFRNKVVTVDGARVKLQIWDTAGQERFRSVTHAYYRDAQALLLLYDITNQSSFDN 120
```

```
Query: 402 IRAWLTEIHEYAQRDVVIMLLGNKADMSSERVIRSEDGETLAREYGVPFLETSAKTGMNV 581
          IRAWLTEIHEYAQRDVVIMLLGNKAD+SSERVIRSEDGETLAREYGVPF+ETSAKTGMNV
Sbjct: 121 IRAWLTEIHEYAQRDVVIMLLGNKADVSSERVIRSEDGETLAREYGVPFMETSAKTGMNV 180
```

```
Query: 582 ELAFLAIAKELKYRAGHQADEPSFQIRDYVESQKKRSSCCSF 710
          ELAFLAIAKELKYRAG Q DEPSFQIRDYVESQKKRSSCCSF+
Sbjct: 181 ELAFLAIAKELKYRAGRQPDEPSFQIRDYVESQKKRSSCCSFV 223 (SEQ ID NO:4)
```

## **Hmmer search results (Pfam):**

Model	Description	Score	E-value	N
PF00071	Ras family	306.9	8.4e-90	1
CE00060	CE00060 rab_ras_like	213.3	3.7e-60	1
PF01142	Uncharacterized protein family UPF0024	2.6	3.4	1

## **Parsed for domains:**

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
CE00060	1/1	31	191 ..	25	193 ..	213.3	3.7e-60
PF01142	1/1	185	201 ..	444	462 .]	2.6	3.4
PF00071	1/1	31	223 .]	1	198 []	306.9	8.4e-90



3151 TTGTGGCAGG AAATGGGCCC CTGCACCCTC GGAGAGGAGG AGCTGCTGTT  
 3201 GGCCAGGCCC CAGGCTGAGG GGGACTGCCT GACCTTGTGT CCCTGCAAAC  
 3251 CAGCTGGGTT GTTTGGCTAG GAGGTGGCCA GGCTAGGCAG CTGTTGTGT  
 3301 TTGGTGAAT CACCGAGCTG GGTGGGTAGC TGGCATCGTT TGCTCAAGGC  
 3351 AGCTGTGATC TGTAAGTAC ACAAAGACTG GCCCTCCCTC CCTCCTTCCT  
 3401 GCTCCAGGGC TGGGACCCAG GAGCCAGGGA GGAGTGCAGG CTCCAGAAAG  
 3451 CTCCTATCCC CCACCCCTTC ATCTGTTCCT TGGCCAAGCG GCATTGGCCG  
 3501 GAGAGTTGGT CCCCAGCCTC CCCGGGCTG CCCCAGGGGA GTGAGTCCAG  
 3551 GACCTCTGA GAAAGCCTGG CAGGAGCTCC TTGGACCAGA CTAGGGGTGA  
 3601 TGTGGCCAC AGGCAGACAG TTCCACCTT GGGCCACTCT TCCCTGGGTC  
 3651 TTAGGTGATT CACCACGATG ATGGGCCCTA GCCATTAACA GACTCTAGAA  
 3701 ATACCTCAA GACATTATCC CTCCTCCTTC TACCCACTAT GGAAACCATG  
 3751 CCACAGAAAG GTTAAGGAAT CTTCTAAAG TCACACAGTA GGCCATTTAC  
 3801 AAATCAAGAC CCATCCTTCA TACCCCTTCT GCTCAGCCAC CCCTGCCTCT  
 3851 CCACCAGAGT TAACTAATGC CAGTACCCCA TGCCCAACA AGGAATGCCT  
 3901 TTGGGTCCA CTGTCAATTT CAGAGCCTCA AAAATAATTC AAACCTAGTC  
 3951 CCTGCTTAAC ACCTAAGCC AGCTGGGAA ATTCCAGCAT  
 4001 TGGATCTAGA CCCCTGTTAT CCAAGATTGG AGAACAGTGG GACAAAGTGC  
 4051 TCCTCTCCAC CATTCCTGCG TGTCCCTGGG GAAGATGAGC AGAGCAGAGC  
 4101 CAGACAGTAA AGGAGAGGGC CACGCCCTC CCACAGGTTA CCTCCTTGGT  
 4151 ACTCCTGCCC GCACTACCCA CAGCAACCCC GGGATGCCGA TCTGCAGCCA  
 4201 CATGTCCCAT GTGGGAGGTT TCTGCTGAAA GAACTTCCAA CTACACATCT  
 4251 CCCCACTCA GTATAAATTT CAACCTTCCC TAATTCATGC AACCTTTTTT  
 4301 TTTTTTTTTT TTTTTTGAGA CAGAGTGTG CTCTGTACC GAGGCTGGAG  
 4351 TTCAGTGATG CAATCTCGGC TCACTGCAAC CTCTACCTCC TGGGTTCAAG  
 4401 CTATTCTCCT GTCTCCGCT CCCAAGTAAC TGGGACTACA GGCGTGTGCC  
 4451 ACCACTCCTG GCTAGTTTTT TGTATTTTTA GTAGAGATGG GGTTCACCT  
 4501 TGTGGTCTAG GCTGGTCTCA AACTCCCAAC TCAGGTGATC CGTCCACTTG  
 4551 GGCACCCAAA ATGNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN  
 4601 NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN  
 4651 NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN  
 4701 NNNNNNNNN NNNNNNNNN TTCAAGTACC AGCCTGGCCA ACATGGTAGA  
 4751 AACCCTGCT CTACTAAAAA TAAAAAATTA GCCAGGCGAG GTGGTGCATG  
 4801 CCTATAATCC CAGCTACTCA GGTAGGCTGA GGCAGGAGAA TCATTTAAAC  
 4851 CTGGGAGGTG GAGGTTGTGG TGAGCCAAGA TCTGCCATT GCACTCCAGC  
 4901 CTGGGCAACA AGAGCAAAAC TCCGTCTCAA AAAAAAAG AAAGAAAGAA  
 4951 AGAAAGAAAC TTCCAAATAA ATGTTGTGAC ACAAAAAA AAACCCAAAC  
 5001 AATATTCATT ATAGAGTATG CAAATGACCA TGCCCCACCC CCAGCAGATT  
 5051 CTGATAGACT CCCTTGGGTG GGAATCCTTG TCCAATATAT TGACACTTCC  
 5101 CTTTCTGTG AGTATAGCCC AGCCATGCG TGTACTCACG AGCGGACGAT  
 5151 GGATGACACA AGTACACAGA GGGACGGAAT CCCTGCATGG TGTGGCTATG  
 5201 GGCAAAATGT GCCACTGTCT AGATTGTGCA AATGTGGTGG TTCTCTGGGG  
 5251 CCACAGAGCA CACTTGGGGA CCTGTTCTAT GTGAGGTCTC AACTCGGGC  
 5301 TCTAGGAAT TGAATGAGGA CAGGAGGGTC AGAGGGAGAG CCTAGGAGGC  
 5351 TGAGCCAAGG AGCGTGGAGA GGAGAGACAG GGTGAAGGTG GCGGCTGGCT  
 5401 TTCTGGAAGC AGGTGGCCTT TGGTGGGTC AGCATTCGTG CCAGCCCCCT  
 5451 CTTCTCTGAT CCTCTCCATG TGTCTCTCTC CTGGAATCCC AGAAGCTGCC  
 5501 CTTGACTCCC CATTAAGTGC CTCTGCCCCC ACCCCCTAGG TGATGCTTCT  
 5551 GGGAGACACA GCGTGGGA AAACATGTTT CCTGATCCAA TTCAAAGACG  
 5601 GGGCTTCCT GTCCGGAACC TTCATAGCCA CCGTCGGCAT AGACTTCAGG  
 5651 GTGAGGTGGC TGCAGGCACT TGCTTCCAGC AGAGAGCCAG GGCTGTGGCT  
 5701 CAGGCATGGG GGGGTGCCCC CCACCTTGCT CACCCTGGCT CCCAGGGACT  
 5751 CCCGAGGCTC ATGCCTGGAG GGCACACAAC CCGTCCCCC AAGACCACAG  
 5801 AGGTGGCCGG GTCAAAGGAG ACTGGGCAAG GTTGGCTCCT TGCCCAACTA  
 5851 TAGGATGCAA AAAAATGAGA CTGAGTCTTC GATTCCAGCT CCATTCTTGG  
 5901 GGGACTTCTC CCAAGCAGAG CAGCCGAGG CACGGCATAA GCTGAATATC  
 5951 TTGGGCCCACA GAGCCCTGCT TCATTGCTCT CCTACCTGGG CCCCTTTGGA  
 6001 AAGGCCCTCA AGGTCAATCA GTCTTTCTGG AGTTCCAGA AAGCACAGCC  
 6051 CTGCACTGGG TTTAAGAGCT GGGCTTGGGC CAGGCATGGT GGCTCTTGCC  
 6101 TGTATTCCCA GCACTTTGGG AGGCCGAAGC GGTGAGATCA CAAGGTCAGG  
 6151 AGTTTGAGAC CAGCCTGGCC AACATGGTGA AACCCCTGCT CTAATAAAAA  
 6201 TACAAAAATT AGCCAGGTGT AGTGGCACGC TCCTGCAGTC CCAGCTACTC  
 6251 GGGAGGCTGA GGCAGGAGAA TCGCTCAAAT CCGGCTGGTG GAGGTTGCAG

FIGURE 3, page 2 of 7

TGAGCTGAGA TCGCGCCACT GCACTCCAGC CTGGGCAACA AAGTGAGACT  
 GCGTCTCAGA AAAAAAAAAA AAAAAAGAGC TGGGCTGGCC ATGTTGGGAG  
 ACAGCAGCTC ACCAGGGACC CTCCCTCTCA CCTTGACGAC TCCATCTTAC  
 AAATCTGCAT CAGGGATGCT AGACGCTGCA CACCTGAAGT GTTCAATAGA  
 GAAAAGGTCT CACCCTGGCA GGTGGGGCTC TACAGCTTCA AGCAGGCAGA  
 AAGCGAACAC TTCCTTCACT AGAGAATTAG TGGGCAGCTA AAGAAAAGGT  
 GCTGCTGCAG ATGTAGCCTC AGGTCCCCAG GATGCAGGCA AACACCCCAT  
 CTCCAGGGGC TCGGTACACAG TCCCAAGGCT AGGCTCCAGG AGAGGGAGAC  
 CGAAGTGGGG AAAGGGCAGG GCCTCCAGCA GCAACCAGCC CTCCAGCCCT  
 GGGCTGCCTG ATCCCTGGAG AGAGCCAGGA TGTTTCTCAG GCTCCTCTTG  
 CCCTGCTGTT GTGAGAAGGC AGTTACAGTC CTCAGAAGGG ACGACTCCAC  
 AGTGGAGGTG TCTGGGTATG GGGTTCCTGC TGCCCTGATG GTATGATCTG  
 GCTGGAGACG GTTCTGGGGC TCACTGCACC CACTCTAGGC CTGGAGAGGG  
 AACAAGAGAG GACGTCTGCA GAGCTGAGGA GCCACATGAC TCCTGCCCTC  
 CCATCCTCTG CCTTTTCTC TTTCAGAACAA AGGTGGTGAC TGTGGATGGC  
 GTGAGAGTGA AGCTGCAGGT GAGACCAGAG GCTGGAGTTG GGGAGGGAGG  
 ATGGAGGACC TGCCCTTCTT TCTCACCCTG AACCACAGGA GGCTGCAGC  
 CCTGCCCTCC GCCTGGGGCA ATTTCTCTGTG GGGCCACCGG GAGGAAATGG  
 CTTTTGTTTA TTTGACATCT GCAGAAAAAG CAGTCCCAG GCACCCTCTC  
 ATCTATGAAC AGCAGCTCCA AATGCCTTCA GACAAGCTTA GCCTCCATCC  
 ATCTCCTCCC CAGTTGCCAG GGCTTTATCT GCTCTTAGGA GATTGGACAT  
 CCCCACCCCC TGAGCTAGGG GAGAGGAGAA GATTCTTTTT TTTCTTTTC  
 TTTCTTTTT TTTTGTAGA TGGAGTCTCG CTCTGTCGCC CAGGCTGGAG  
 TGCAGTGGCA CAATCTCGGC TCACTGCAAC CTCTGCCTCC CAGGTTAAG  
 AGATTCTCCT GCCTCAGCCT CCTGAGTAGC TGAGACTACA GGTGCATGCC  
 ACCACACCTG GCTAATTTTT TGTATTTTGA GTAGAGACGG GGTTCCTG  
 TGTTAGCCAG GATGGTCTGG ATCTCCTGAC CTCGTGATCC GCCTGCCTCG  
 GCCTCCCAAA GTGCTGGGAT TACAGGTGTA AGCCACCGCG CTCGGCTGAG  
 GAGATGATTT TGAACGAGCT TGAGAAATCA GTAAGTCTA CTGTCCAGGT  
 CATTGGATGC TAAGGGGCTC ATGAGAACCT AAAGAAGAAA ACAGCCCCAC  
 CTTCCACAG ATATCTCATA CAACAAAGCA GGCTGCTCC ACCCAGCACA  
 TTCCTTGAC CTGCCTCCTT CTGACCATT CTCCATCCCA TCCCTTCCCA  
 GATCTGGGAC ACCGCTGGGC AGGAACGGTT CCGAAGCGTC ACCCATGCTT  
 ATTACAGAGA TGCTCAGGGT GAGTCCCTCG CACCTCCAA CCCCTACCCC  
 AGCCCCCTGG TAGCATCCGT GCTGCTGCCT AAGTCCCCTC TGTGATCCTC  
 TCCCCTCCAG CCTTGCTTCT GCTGTATGAC ATCACCACAA AATCTTCTTT  
 CGACAACATC AGGGTAGGTC CTCCCTTCCC CTGACTCCCA CCCATAAGCA  
 GCCAAGGCAA GGTCTATGCA GGCTGGGGTT GCTTCTGCC CTGTGGAAG  
 CGGGTGGAGC GTGGAGTCTT CCTGCCCTCT GAAAAACACC TACTTGTGAC  
 TCAGAAAGTCA TATCTGCTGC TTTGTATTTG GTGGCCATGT GGGCATGAAG  
 GCCAAGCAGG CTGTTGTGAC CCTGTGCCAC CTGCATAGCC CTCACTGTGA  
 TTCACGAGTG TGTTTCGTGA CAAAGTGTTT AGAACAGCCC CCACTCCACC  
 CTGGATAATT ATCCACAGAG ACCAAGGGAA AAACACAACC AGAAAAGTCC  
 ACACATACAT CAGGGCAAG TTGCAAGAAA GTGACTCAGT CAGACAGAGT  
 GAGTGGTTGT ATCTCACAA CCAAATATT ATAGAGACAA AAATTTGATA  
 AATTCAAGCA CCAATTTTGT TCACGACATT GTATAGGTTT CATGAATCCC  
 CTGACCTCAA GGACAGTTTG CTGATAAGCA AACTAGGAGA ATAAAACGTT  
 TATATAGAAA GAGGAAAATC CATGGCACTC ATACTCCTAC CTCCAACCCC  
 ATGCTCATGG CAGACATCAC TAATCAATCA CAGTACTTTT GATCACTGAA  
 ACCCTTATGT GGTCTTAGAA TCTTTAACAG GACACTCCAA GAAATCACTG  
 CTGACAGCCA ACTGATTTGT GAGATAAGGT CTCCATGCAT CTGGATCTTC  
 CATAGAACTG ATAGTTGCAC AGCATAAAAT GGTGAGGGTG GGGCCATTGT  
 GGGTTGAGCC ACCAAGGAAG GCCATCCAGG CCTGGATGGG CCAGAACAAA  
 GGTACAGATG AGAGAACGCA CAGGGTATCG TGTTCAAGGT AGTGAGTAAC  
 TGAGGATAGT CAAACGGAGC AGAAGAAGAA AGGGGCAGCA GGAGGAAGAG  
 AATGCCAGTC TCGCACGCC TCTCCACAG GCCTGGCTCA CTGAGATTCA  
 TGAGTATGCC CAGAGGACG TGGTGATCAT GCTGCTAGGC AACAAAGTGA  
 GTGGCTCCGG GGCAGGGTCA GCCCAGCCCT GCACTTCCTC AGCCCTAGCC  
 GGCCCCATAA CCACCCAAGA ACAGTTATCT AGGCATCCTT CCTGAAAAGG  
 ACTCTGACG CTCCAGCTCA GGGGTGAGC ATATCTGGAG GCTCTGCCC  
 ATCCCATCTG CCCCTTCCAG GGAAAGTCCA AGTTGTTGCC TGAGAAATCA  
 AGGGGTGCCC AGTTCTCAGC CCCCATTAGA GCAGAGTGAA CAGGGTCCCA  
 GGTCAGGGGC TAAGAGTGCA AAGGGTTAGC CCCAACTGCT GTCCTATTCC

FIGURE 3, page 3 of 7

9451 AAGACCCTTT ACCAAAGGTG AGATCCCAGA GCTGGGAGCT AACTGGGGCA  
9501 GAAACCCTGG CCCCAGGCCA ATCACACCTG CCTGCAGTCC CTTGGGGCCAC  
9551 CAGCAGAGGG CAGGCAACGC CTGCTTCTGG GGCAAAATAT GGGCCCCGTG  
9601 GGGCGGAGGC CTCCTTCCCC AGAGTGACCC ATTTGGGCTT GACAGGCGGA  
9651 TATGAGCAGC GAAAGAGTGA TCCGTTCCGA AGACGGAGAG ACCTTGGCCA  
9701 GGGTAAGTGA TTGTCTGTGG GACAGGGTGA AGGGTGGGGG CAACCCGACG  
9751 CTGGCCCTGA GGACACTCTC TCCCGGGCAG GAGTACGGTG TTCCCTTCCT  
9801 GGAGACCAGC GCCAAGACTG GCATGAATGT GGAGTTAGCC TTTCTGGCCA  
9851 TCGCCAAGTG AGAGCTGGGC AGGGAAGGGA AGTGTGCGGG GCAGGGCGGC  
9901 ACACCTCCAGG AATCCAGTAG GGCCCGGCC CTGGCCAGC CCCTGGACAC  
9951 ACCTGCATTC TGCAGGCTGA GGTCCATTTG CTCTGGGAGC ACTGGGCCAC  
10001 TGGGAGAGGG GAGGGGGCGG CTCAGCTCCT CACCCAGCC CAGCCAGCC  
10051 CAGCCAGCC CATTGTCTCT TCTCAAGGG AACTGAAATA CCGGGCCGGG  
10101 CATCAGGCGG ATGAGCCAG CTTCAGATC CGAGACTATG TAGAGTCCCA  
10151 GAAGAAGCGC TCCAGCTGCT GCTCCTTCAT GTGAATCCCA GGGGGCAGAG  
10201 AGGAGGCTCT GGAGGCACAC AGGATGCAGC CTTCCCCCTC CCAGGCCTGG  
10251 CTTATTCCAA GAGGCTGAGC CAATGGGGAG AAAGATGGAG GACTCACTGC  
10301 ACAGCCGCTT CCTAGCAGGG AGCTATACTC CAACTCCTAC TTGAGTTCTT  
10351 GCGGTCTCCC CGCATCCACA GGGAGGGTAA AACACTTAGC TTTTATTTTA  
10401 ATAGTACATA ATTTAATACC AAAAAAGCG CCTGGATCCC CAAAAACCG  
10451 AGGCTGGGAG CTAGTGCCCC TTTTGCTTTC TAGGACTTGG GGGGCCGGCC  
10501 CTCCCTCCTA AGCATAACAA AGGTGGTGTT GCTCCAGCTC AGCCCCAGGG  
10551 GACACAGATG CACTTTGGGG GTGAGGGCAG GTAATGACTC CATCGCACCC  
10601 TCAGTTTCAAG TGGACAGAGG CTCAGGTGAC CCCAGCCTTC ACTGTCTCCC  
10651 GCTCTCCAGG AGCTTATCTT CGCCCCATCT CCCAAATAAG TGGGCCCTTG  
10701 TGCTGTGAGG AAGACCAAG CCTCAGGGAA GATAAGAGAT ATGGAGATGG  
10751 GAGGGGGAGG ACAAGGGGCA GAGAGTAGGG TCTAGCTGGC TATCTCTGGC  
10801 CTTACTAACA CCCCCCTGGA GGCATGCCCC TTTTCTCCAG CACACAAGCA  
10851 CATTGGGGGA CCTGGAATAA TTGGTTCCAG GCTCCTGTTC TCTGGACTTC  
10901 AGATCTGGG GGAGCCCCCTC CCCCCCTGA ATCCCTGGCT TAGCTACCTT  
10951 CCTGCCTGTG CACCTAAAAA CCTCAGGTCA GAACTAGGAA AAGAGTTTTG  
11001 TTTTATTTTT TTTGAAATGG AGTCTCGTTC TGTGCGCCAG GCTGAGGTGC  
11051 AGTAGTGCAA TCTCCGCTCA CTACAACCTC CACTCCCTGG GGCTCAAGCG  
11101 ATCCTCCAC CTCAGCGGCC GAAGTAGCTG GGACTATAGG TGTGTACCAT  
11151 CACACCTGGC TAATTTTTGT ATTTTTTGTA GACACAGGGT TTCGCCATGT  
11201 TGCCCAAGGT GGTCTTGAAT TCCTGAGCTC AAGCAACCTG CCGGCCTCGG  
11251 CCTCCCAAG TACTGGGATT ACACGCAGAA GGCACCATG CCAGGCTAGA  
11301 TGTGTCTTAT CCAATCCTT TGGCAGGCAT GCAGCTCCAC AGGCGATTTC  
11351 TTCAAGCAGC TGAAGTGTTC AGCCCTCCTG GGTAAAGAGC CAGATAAGGA  
11401 GAAATCCCTT TCCTAGGTTT GGAATGTGTT GTGAAAAAA AGAGAAATCC  
11451 CTGGCTCCTG GAGCTGGTGG GAGACAAGAT TAAGCAAACC TCCCTGACA  
11501 TGTATCCCTT TGACCCCAAG CTCTGCCTCC TCCCTGACCA CCCATGCCCT  
11551 TTCCTTTAAT TTCTCAAACA GATACAGGG CCTAAACTGC TTTACCTCCC  
11601 CTCCTACTGA GTCAGGTAG GTGGTGGGAG GTCACCCATT TCCGAGTTAA  
11651 ACCAATGCAA TATGAGTAAA ACAAAGTCAT GTGGGTATGT CTGGGGTAGA  
11701 GAGAGGGGTA GCAAGTTCAT GTGTCTCCT TGGTCACATA TCTCCCAAAG  
11751 CTCTGATCCC TGCCATGGGA AGTGGACAGG AAACATGAGG TCATGACCTG  
11801 CAGGCATCTT TACTGCAGCT CTGCCGGCCT GGAGGGGGAG AGGGGGAGGA  
11851 AGAAGTATGC GCTGCACATT TCTGAGGCTA CTGCATTTGC TTTCAAGGCA  
11901 GAAATCTTGC TCTGAGCAGT CAGCGGCTCC AGTTTGGGCC CGATAAGGAA  
11951 GTTCTCCGTG GCCTCCCTCA GGCAGAGCAG GGAGGAGGCT GACATTGCCA  
12001 GTCTCTTCTG GGGCCCAAGG CAGGTTGCAG GAGATCCAAT CCCATAGACA  
12051 GCTCTGGGCC TCTTGCAATT GAGTTTTTCA GAATTAAACT GCAGTATTTT  
12101 GGAAAGCACA TCCTGTCCAC TGTTTCTTTG AAGTGAGTGG GGGGGGGGGG  
12151 TCTTGTGAA GGAATTGTCA TTTACTGCCA AAATCATTC ATCTCCTTC  
12201 CTCAGTGTCT GTCCTCAGAT GGTGAGCTCC CCGCTCAACA GACTGTCTCC  
12251 CGCCTCTGTG ACCAGCCTCT CTTTGGCAAG AGGGAGCTAG AAGGCTTTAC  
12301 AGTCCTAATC ATTTTCTGT TGGAAAAAA AAAAAAAAC CAAGGCTCCT  
12351 TTCCCTGTGG CGTGTACCCA GAGGTTGATT ACCTGAGTCT GTCTGCCTC  
12401 TCCCCACCCC ACCTCCCTAG CCAAACGCTG CTGCCAAAGC CCACGCTATT  
12451 GCCCTAGATG GCCTGTCTTC AGCGGGCTGC CCCTCGAGGT CCCAGGCTCT  
12501 CCGCGGAGCC CTCACCTTCC CAGCAGGGAT CAGAACCTGC ACTCCTCTAT  
12551 GCGAGTCTTG GGACAGCACA AAGTGGATTA GGTTAGGGT TCCACAAAC

FIGURE 3, page 4 of 7



12601 GGAAAAATGT TATTCAAACA ACTCTGTAGG GTCCGAGGAG GCCCTCCGTC  
 12651 TTAATTCTCG AGACTGACCG GCCCTCGCTG CCCCAGACGG GAGCAGTTGC  
 12701 CCCGGCAACA GCCGCTCCCT CTCAACTGGA GCTGCACCCA GGCTTTGGCT  
 12751 AAAGGCTGTT AAAACGTTGG CCAGGTGCGG AGGCTCACGT CTGTAATCCC  
 12801 AGGGCGGATC ACCTGAGGTC AGGAGTTTGA AACCATCCTG GCCAACATGG  
 12851 CGAAATTTTCG TCTCTACTAA AAATACAAAA ATTAGCGGGG CGTGGTGGTG  
 12901 CGCGCCTGTA ACCCCAGCTG CTCGGGAGGC TGAGGCAGGG GAATCGCTTG  
 12951 AACC CGGGAG GCGGAGGTTG CAGTGATCCG AGATCGCGCC ACGGCAGTCC  
 13001 AGCCTGGGCG ACAGAGCGAG ACTCCGTCTC AAAAAAAAAA AAAAAAGTTA  
 13051 GGGTCCTTTA CCCGAGGGCC GGCTTTCCTC ACTCCCCGCC ACAGGTAGGG  
 13101 GAAACCAGGC CGGAGCCGGC GGGCCACCC GCCCAGAACC GGGAATTCGG  
 13151 CGAGCCCCGC CCCTGCCACC CCAGCGCCGG CC (SEQ ID NO:3)

#### FEATURES:

Start: 2042  
 Exon: 2042-2134  
 Intron: 2135-5539  
 Exon: 5540-5650  
 Intron: 5651-7026  
 Exon: 7027-7068  
 Intron: 7069-7901  
 Exon: 7902-7968  
 Intron: 7969-8060  
 Exon: 8061-8113  
 Intron: 8114-9080  
 Exon: 9081-9146  
 Intron: 9147-9645  
 Exon: 9646-9702  
 Intron: 9703-9780  
 Exon: 9781-9857  
 Intron: 9858-10078  
 Exon: 10079-10181  
 Stop: 10182

#### CHROMOSOME MAP POSITION:

Chromosome # 17

#### ALLELIC VARIANTS (SNPs):

DNA			
Position	Major	Minor	Domain
4259	C	T	Intron
4325	G	T	Intron
4348	G	A	Intron
4924	G	A	Intron
4983	-	A	Intron
6710	A	G	Intron
8624	A	G	Intron
8661	G	A	Intron
11754	T	C	Beyond ORF(3')
11836	A	G	Beyond ORF(3')

Context:

DNA  
Position  
4259

ACCCATTAAGCCACCTAACCAGCAGCTGGGAAATTCCAGCATTGGATCTAGACCCCTGTT  
ATCCAAGATTGGAGAACAGTGGGACAAAGTGCTCCTCTCCACCATTCTGCGTGTCCCTG  
GGGAAGATGAGCAGAGCAGAGCCAGACAGTAAAGGAGAGGGCCACGCCCCCTCCACAGGT  
TACCTCCTTGGTACTCCTGCCCCGCACTACCCACAGCAACCCCGGGATGCCGATCTGCAGC  
CACATGTCCCATGTGGGAGGTTTCTGCTGAAAGAACTTCCAACCTACACATCTCCCCACTT  
[C, T]  
AGTATAAAATTTCAACCTTCCCTAATTCATGCAACCTTTTTTTTTTTTTTTTTTTTTTTGAG  
ACAGAGTGTGCTCTGTCCACGAGGCTGGAGTTCAGTGATGCAATCTCGGCTCACTGCAA  
CCTCTACCTCCTGGGTTCAGCTATTCTCCTGTCTCCGCCTCCCAAGTAACTGGGACTAC  
AGGCGTGTGCCACCACTCCTGGCTAGTTTTTTGTATTTTAGTAGAGATGGGGTTTCACC  
TTGTTGGTCAGGCTGGTCTCAAACCTCCCAACTCAGGTGATCCGTCCACTTGGGCACCCAA

4325 GATTGGAGAACAGTGGGACAAAGTGCTCCTCTCCACCATTCTGCGTGTCCCTGGGGAAG  
ATGAGCAGAGCAGAGCCAGACAGTAAAGGAGAGGGCCACGCCCCCTCCACAGGTTACCTC  
CTTGGTACTCCTGCCCCGCACTACCCACAGCAACCCCGGGATGCCGATCTGCAGCCACATG  
TCCCATGTGGGAGGTTTCTGCTGAAAGAACTTCCAACCTACACATCTCCCCACTTCAGTAT  
AAATTTCAACCTTCCCTAATTCATGCAACCTTTTTTTTTTTTTTTTTTTTTTTGAGACAGA  
[G, T]  
TGTCGCTCTGTACCCGAGGCTGGAGTTCAGTGATGCAATCTCGGCTCACTGCAACCTCTA  
CCTCCTGGGTTCAAGCTATTCTCCTGTCTCCGCCTCCCAAGTAACTGGGACTACAGGCGT  
GTGCCACCACTCCTGGCTAGTTTTTTGTATTTTAGTAGAGATGGGGTTTCACCTTGTG  
GTCAGGCTGGTCTCAAACCTCCCAACTCAGGTGATCCGTCCACTTGGGCACCCAAAATG

4348 TGCTCCTCTCCACCATTCTGCGTGTCCCTGGGGAAGATGAGCAGAGCAGAGCCAGACAG  
TAAAGGAGAGGGCCACGCCCCCTCCACAGGTTACCTCCTTGGTACTCCTGCCCCGCACTAC  
CCACAGCAACCCCGGGATGCCGATCTGCAGCCACATGTCCCATGTGGGAGGTTCTGCTG  
AAAGAACTTCCAACCTACACATCTCCCCACTTCAGTATAAATTTCAACCTTCCCTAATTCA  
TGCAACCTTTTTTTTTTTTTTTTTTTTTTTTGGAGACAGAGTGTGCTCTGTACCCGAGGCTG  
[G, A]  
AGTTCAGTGATGCAATCTCGGCTCACTGCAACCTCTACCTCCTGGGTTCAAGCTATTCTC  
CTGTCTCCGCCTCCCAAGTAACTGGGACTACAGGCGTGTGCCACCACTCCTGGCTAGTTT  
TTTGTATTTTAGTAGAGATGGGGTTTCACCTTGTGCTCAGGCTGGTCTCAAACCTCCCA  
ACTCAGGTGATCCGTCCACTTGGGCACCCAAAATG

4924 TTCAAGTACCAGCCTGGCCAACATGGTAGAAACCCCGTCTCTACTAAAAATAAAAAATTA  
GCCAGGCGAGGTGGTGCATGCCTATAATCCCAGCTACTCAGGTAGGCTGAGGCAGGAGAA  
TCATTTAAACCTGGGAGGTGGAGGTTGTGGTGAGCCAAGATCTCGCCATTGCACTCCAGC  
CTGGGCAACAAGAGCAAACTCC  
[G, A]  
TCTCAAAAAAAAAAGAAAGAAAGAAAGAAAGAACTTCCAAATAAATGTTGTGACACAA  
AAAAAAAAAACCCAAACAATATTCAATTATAGAGTATGCAATGACCATGCCCCACCCCAAG  
CAGATTCTGATAGACTCCCTTGGGTGGGAATCCTTGTCCAATATATTGACACTTCCCTTT  
CCTGTCAGTATAGCCAGCCCATGCGTGTACTCACGAGCGGACGATGGATGACACAAGTA  
CACAGAGGGACGGAATCCCTGCATGGTGTGGCTATGGGCAATGTGGCCACTGTCTAGAT

4983 TTCAAGTACCAGCCTGGCCAACATGGTAGAAACCCCGTCTCTACTAAAAATAAAAAATTA  
GCCAGGCGAGGTGGTGCATGCCTATAATCCCAGCTACTCAGGTAGGCTGAGGCAGGAGAA  
TCATTTAAACCTGGGAGGTGGAGGTTGTGGTGAGCCAAGATCTCGCCATTGCACTCCAGC  
CTGGGCAACAAGAGCAAACTCCGTCTCAAAAAAAAAAGAAAGAAAGAAAGAAAGAAAC  
TTCCAAATAAATGTTGTGACAC  
[-, A]  
AAAAAAAAAACCCAAACAATATTCAATTATAGAGTATGCAATGACCATGCCCCACCCCA  
GCAGATTCTGATAGACTCCCTTGGGTGGGAATCCTTGTCCAATATATTGACACTTCCCTT  
TCCTGTCAGTATAGCCAGCCCATGCGTGTACTCACGAGCGGACGATGGATGACACAAGT  
ACACAGAGGGACGGAATCCCTGCATGGTGTGGCTATGGGCAATGTGGCCACTGTCTAGA  
TTGTGCAATGTGGTGGTTCTCTGGGGCCACAGAGCACACTTGGGGACCTGTTTCATGGTG

6710 CACCAGGGACCCTCCCTCTCACCTTGACGACTCCATCTTACAAATCTGCATCAGGGATGC  
TAGACGCTGCACACCTGAAGTGTTC AATAGAGAAAAGTCTCACCCTGGCAGGTGGGGCT

FIGURE 3, page 6 of 7

**THE** **WORLD'S** **LARGEST** **BOOKSTORE**

11836 GACCACCCATGCCCTTTCTTTAACTTCTCAAACAGATACCAGGGCCTAAACTGCTTTAC  
CTCCCCTCCTACTGAGTCAGGTTAGGTGGTGGGAGGTCACCCATTTCGAGTTAAACCA  
TGCAATATGAGTAAACAAAGTCATGTGGGTATGTCTGGGGTAGAGAGAGGGGTAGCAAG  
TTCATGTGTCTCTCTTGGTCACATATCTCCAAAGCTCTGATCCCTGCCATGGGAAGTGG  
ACAGGAAACATGAGGTCATGACCTGCAGGCATCTTACTGCAGCTCTGCCGGCTGGAGG  
[A, G]  
GGAGAGGGGGAGGAAGAAGTATGCGCTGCACATTTCTGAGGCTACTGCATTTGCTTTCAA  
GGCAGAAATCTTGCTCTGAGCAGTCAGCGGCTCCAGTTTGGGCCCCGATAAGGAAGTTCTC  
CGTGGCCTCCCTCAGGCAGGACAGGGAGGAGGCTGACATTGCCAGTCTCTTCTGGGGCCC  
AAGCAGGTTGACAGAGATCCATCCCATAGACAGCTCTGGGCCCTCTGCATTTGAGTTT  
TTCAGAATTTAACTGCAGTATTTTGGAAAGCACATCCTGTCCACTGTTTCTTTGAAGTGA